A statistical approach to central monitoring of data quality in clinical trials

David Venet, Erik Doffagne, Tomasz Burzykowski, François Beckers,
Yves Tellier, Eric Genevois-Marlin, Ursula Becker, Valerie Bee,
Veronique Wilson, Catherine Legrand and Marc Buyse

Clinical Trials published online 8 June 2012
DOI: 10.1177/1740774512447898

The online version of this article can be found at:
http://ctj.sagepub.com/content/early/2012/06/08/1740774512447898

A statistical approach to central monitoring of data quality in clinical trials

David Venet [a,b], Erik Doffagne [a], Tomasz Burzykowski [a,c], François Beckers [d], Yves Tellier [d], Eric Genevois-Marlin [e], Ursula Becker [f], Valerie Bee [g], Veronique Wilson [g], Catherine Legrand [h], Marc Buyse [c,k]

Background
Classical monitoring approaches rely on extensive on-site visits and source data verification. These activities are associated with high cost and a limited contribution to data quality. Central statistical monitoring is of particular interest to address these shortcomings.
Purpose
This article outlines the principles of central statistical monitoring and the challenges of implementing it in actual trials.
Methods
A statistical approach to central monitoring is based on a large number of statistical tests performed on all variables collected in the database, in order to identify centers that differ from the others. The tests generate a high-dimensional matrix of p-values, which can be analyzed by statistical methods and bio-informatic tools to identify extreme centers.
Results
Results from actual trials are provided to illustrate typical findings that can be expected from a central statistical monitoring approach, which detects abnormal patterns that were not (or could not have been) detected by on-site monitoring.
Limitations
Central statistical monitoring can only address problems present in the data. Important aspects of trial conduct such as a lack of informed consent documentation, for instance, require other approaches. The results provided here are empirical examples from a limited number of studies.
Conclusions
Central statistical monitoring can both optimize on-site monitoring and improve data quality, and as such provides a cost-effective way of meeting regulatory requirements for clinical trials.

[a] *International Drug Development Institute (IDDI), Louvain-la-Neuve, Belgium*
[b] *Institut de Recherches Interdisciplinaires et de Développements en Intelligence Artificielle (IRIDIA), Free University of Brussels, Brussels, Belgium*
[c] *Interuniversity Institute for Biostatistics and statistical Bioinformatics (I-BioStat), Hasselt University, Diepenbeek, Belgium*
[d] *GlaxoSmithKline Biologicals, Wavre, Belgium*
[e] *Sanofi-Aventis R&D Biostatistics and Programming, Bridgewater, NJ, USA*
[f] *F. Hoffmann-LaRoche Ltd., Basel, Switzerland*
[g] *Translational Research In Oncology (TRIO), Paris, France*
[h] *Institute of Statistics, Biostatistics and Actuarial Sciences (ISBA), Université Catholique de Louvain, Louvain-la-Neuve, Belgium*
[k] *International Drug Development Institute (IDDI) Inc., Houston, TX, USA*

Author for correspondence:
Marc Buyse, ScD IDDI Inc., 363 N. Sam Houston Parkway East, Suite 1100, Houston, TX 77060, USA, Email: marc.buyse@iddi.com

INTRODUCTION

Over the last decade, clinical research costs have sky-rocketed while new drug approvals have decreased by one third [1]. At the current pace of increase in costs, adequately sized clinical trials will either become infeasible, or, because such costs will have to be reflected in the price of new drugs, they will eventually cause an intolerable burden on health care systems. Eisenstein et al [1] have studied typical trial costs and suggest different options to reduce some of these costs without compromising the scientific validity of the trials. The greatest potential savings lie in the labor-intensive activities such as on-site monitoring, which can represent as much as 30% of the total budget in large global clinical trials [2]. It is therefore not surprising that the current practice of performing intensive on-site monitoring is coming into question. Recently, pragmatic, risk-based approaches have been proposed that improve the cost-effectiveness ratio without compromising the quality and integrity of clinical trials. A recent draft guidance of the U.S. Food and Drug Administration (FDA) [3] reflects this trend and states unequivocally: "*FDA encourages greater reliance on centralized monitoring practices than has been the case historically, with correspondingly less emphasis on on-site monitoring*". A recent reflection paper of the European Medicines Agency (EMA) [4] expresses a similar view (footnotes added for clarity): "*Adaptations to conventional GCP[1] methods, for example, adaptation of on-site monitoring visits, sample/focused SDV[2], new central monitoring processes etc., subject to appropriate metrics being captured to determine when/if escalation in monitoring would be appropriate.*" These texts establish the importance of central monitoring as a cost-effective way of ensuring data quality in clinical trials.

In this paper, we first discuss monitoring practices, and current experiments aimed at validating optimized monitoring strategies. We then describe the underlying principles of a statistical approach to central monitoring. We illustrate typical findings obtained with this approach in actual trials from different therapeutic areas. We conclude by a discussion of the potential role of central statistical monitoring and argue that it can both optimize on-site monitoring and improve the quality of clinical trial data.

[1] *Good Clinical Practice*
[2] *Source Data Verification*

BACKGROUND

CLINICAL TRIAL MONITORING

Clinical trial sponsors are required to set up appropriate measures to monitor the conduct of the trial. The aim of monitoring is to ensure the patients' well being, compliance with the approved protocol and regulatory requirements, and data accuracy and completeness [5]. We shall focus our attention solely on the latter aspect of monitoring, leaving aside its other purposes that are arguably more important, such as checking the patient's informed consent forms and training the local staff (usually at the beginning, but also during the study if and when deviations from the study protocol are observed).

Baigent et al [6] draw a useful distinction between three types of trial monitoring: oversight by trial committees, on-site monitoring, and central statistical monitoring. They argue that the three types of monitoring are useful in their own right to guarantee the quality of the trial data and the validity of the trial results. Oversight by trial committees is especially useful to prevent / detect errors in the trial design and interpretation of the results. On-site monitoring is especially useful to prevent / detect procedural errors in the trial conduct at participating centers (e.g., informed consent signed by the patients or legally acceptable representative). Statistical monitoring is especially useful to detect data errors, whether due to faulty equipments, negligence or fraud.

CURRENT MONITORING PRACTICES

The landscape of current monitoring practices has recently been studied in a survey [7] targeting different types of organizations: academic groups, CROs (Contract Research Organizations), pharmaceutical companies and device companies. More than 80% of trials regardless of type of organization included on-site visits with source data verification (SDV) which consists of comparing information recorded in the CRF (Case Report Form) with the corresponding source documents. According to this survey, SDV tends to be done systematically on key variables (e.g. consent, eligibility, serious adverse events and primary outcome), but less so on the secondary variables (e.g. secondary outcomes or non-serious adverse events). SDV detects discrepancies due to transcription errors from source documentation to CRF, but it may miss errors present in the source documents. It has therefore been argued that the contribution of SDV to data quality is minimal, with full (100%) verification of all source data being

particularly cost-ineffective [8]. The survey [7] also revealed that few respondents use a centralized monitoring approach to guide on-site visits. In addition to high costs, the travel-related intensive on-site monitoring also has a non-negligible ecological footprint. It has been estimated that travel by the trial team generates around 20% of the total carbon emission generated by a clinical trial [9].

Extensive monitoring with 100% SDV is no longer favored, even in regulatory trials, and is being progressively replaced by "reduced monitoring", which consists of controlling only a random sample of data. The random sampling can be performed at various levels: country, centers within countries, patients within centers, visits within patients, CRF pages within visits, etc. Reduced monitoring is usually adapted to the risk associated with the experimental procedure. For instance, a trial involving innocuous procedures or well-known treatments could involve far less monitoring than a trial involving invasive procedures or experimental new drugs. Another option is "targeted" monitoring (also known as "adaptive" or "triggered" monitoring), where the intensity and frequency of on-site monitoring is triggered by key performance or risk indicators. These indicators typically focus on critical aspects of trial conduct: accrual performance (e.g. actual accrual rate compared with projected accrual rate, accrual patterns over time), protocol adherence (e.g. percentage of protocol deviations, percentage of dropouts), treatment compliance (e.g. percentage of dose reductions or delays), safety reporting (e.g. percentage of adverse events and serious adverse events reported), and data management (e.g. percentage of overdue forms, query rate, query resolution time). Some authors estimate that targeted monitoring can lead to a reduction of the number of visits by up to 25% compared to traditional approaches [10].

It seems somewhat paradoxical that statistical theory, which is so central to the design and analysis of clinical trials, has not been put to use to help optimize monitoring activities, even though the potential of statistics to uncover fraud in multicenter trials has received some attention [11,12]. In recent years, attention has shifted from the detection of fraud to central statistical monitoring as a tool to detect abnormal patterns in the data and, as such, to help focus monitoring activities on centers where they appear to be most needed [6].

EVALUATION OF MONITORING STRATEGIES

In view of the rising costs of monitoring, several academic groups have proposed formal studies to evaluate the performance of different monitoring strategies. The impact of on-site monitoring on data quantity and quality was evaluated formally in a prospective study embedded in a randomized trial comparing two chemotherapy regimens for patients with breast cancer. In that study, participating centers were randomized to either undergo systematic on-site visits, or no visits at all. Although the study was prematurely stopped and did not fully answer the question it had set out to address, it found no difference in data quality between centers that had undergone an initiation visit, and those that had not [13]. At least two other similar initiatives, driven by academic groups in France and Germany, are currently ongoing. The first one (OPTIMON, for OPTImisation of MONitoring) [14] is a prospective, randomized, non-inferiority trial covering various treatments and indications. The participating centers are randomized to classical intensive monitoring (full SDV, frequent visits) or "light" monitoring, which is modulated according to the risk of the intervention to the patient. The primary outcome measures are the proportion of patients without informed consent problems, the number of serious adverse events reported, errors in the eligibility criteria, and the primary endpoint specific to each study (e.g. tumor shrinkage in oncology). Another initiative (ADAMON, for ADAptiertes MONitoring) [15] also investigates the performance of a risk-adapted monitoring strategy against classical, extensive monitoring. Here too, the participating centers are randomized to either monitoring strategy. At the end of the trial on-site audits are performed to assess the outcome measure of interest, which is the occurrence of serious GCP violations. These initiatives are much needed to provide quantitative evidence of the impact on data quality that can be expected of different monitoring strategies.

METHODS

SOURCES OF DATA ERRORS IN CLINICAL TRIALS

It is useful, before discussing methods to discover errors in clinical trial data, to describe the various ways in which data errors can occur, for different statistical techniques may be required for different types of errors. Note that we focus here on data errors, at the exclusion of other important types of errors that have been reviewed elsewhere (such as design errors,

procedural errors and analytical errors) [6]. It is convenient, for the sake of this limited discussion, to classify data errors into four main classes:

1. **Completely unintentional data errors**, such as data generated by a wrongly calibrated or imprecise instrument. These errors can typically be discovered through a shift in the distribution of values, or an abnormally large variability in this distribution.

2. **Data errors resulting from carelessness**, such as data incorrectly copied from source documents to the CRF, or from the CRF to a database. These errors may be hard to detect but are generally innocuous if their frequency is low. Missing data also fall in this category; these may be damaging to the trial results in so far as it is generally unreasonable to assume that such data are missing at random (let alone completely at random). Fortunately, missing data can be detected through frequency comparisons, with outlying centers (those with too much or too little missing data) being flagged.

3. **Fabricated data**, such as missing or outlying values that are replaced by plausible values, e.g. through interpolation from adjacent values. These errors are typically discovered through an abnormally small variability in the distribution of values, through multivariate tests, or through tests for the similarity of patterns in repeated measures. Extreme cases of data fabrication have been documented, in which entire patients were either invented or extracted from a database of real patients not entered in the trial [16].

4. **Data falsified to reach a desired objective**, e.g. to make a patient eligible, or to show a treatment effect. These errors, when present, can have a devastating effect on the trial credibility, especially if they are aimed at magnifying treatment effects. They can be detected through comparisons of distributions or through center by treatment interactions.

The last two types of errors can be considered fraud and, as such, may need to receive more attention [11]. Indeed, if the fraud is demonstrated, it may not be sufficient to correct or discard problematic data. The validity of the whole trial may come into question, and further actions will eventually be taken against the person responsible for the fraud. Other types of data errors may be important to detect as well, to the extent that they may impact the outcome of the trial. Errors that are unintentional or result from carelessness, that are not too common and are randomly distributed between the randomized groups, have little impact on the trial results, and as such are less important to detect.

PRINCIPLES OF STATISTICAL MONITORING

The second European Stroke Prevention Study (ESPS2) [16] provides a remarkable example of the effectiveness of statistical monitoring as compared with on-site visits for the detection of abnormal data patterns. ESPS2 was a randomized trial of aspirin and dipyridamole in patients with transient ischaemic attack or stroke. The study enrolled 6602 patients at 59 centers, plus 438 patients at a center that eventually had to be excluded from all analyses. "*Fraud or misconduct at the center concerned was considered a possibility early in the recruitment. Despite intensive monitoring this could not be proven one way or the other and external audit was brought in. The audit also failed to establish guilt or innocence…*" [16] In the end, the center was excluded on the grounds that the distribution of dipyridamole and aspirin plasma concentrations differed significantly in the suspect center as compared to all other centers, and was incompatible with the drug administration required by the study protocol. This case exemplifies a situation in which even the most careful on-site review cannot uncover unusual data patterns that are readily detected even by simple statistical methods to compare distributions of continuous variables [11,12]. In less extreme cases, more sophisticated statistical methods will be required. We now turn to general principles underlying statistical monitoring of clinical trial data.

First, statistical monitoring relies on the highly structured nature of clinical data, since the same protocol is implemented identically in all participating centers, where data are collected using the same CRF [11-12]. Hence the same hierarchical data structure is used throughout the trial, with variables or items grouped by CRF page (or screen when electronic data capture is used), CRF pages or screens grouped by visit, visits grouped by patient, patients grouped by investigator, investigators grouped by center, centers grouped by country, and countries grouped by geographical region. When the trial is randomized, the group allocated by randomization provides another design feature that allows for specific statistical tests to be performed, because baseline variables are not

expected to differ between the randomized groups (but through the play of chance), while outcome variables are expected to differ about equally in all centers (but through the play of chance) if the treatments under investigation have a true effect. Abnormal trends and patterns in the data can be detected by comparing the distribution of all variables in each center against all other centers. Similar comparisons can also be made between other units of analysis, if the structure of the trial warrants it. Such comparisons can be performed either one variable at a time in a univariate fashion or with several variables, taking into account the multivariate structure of the data, or using longitudinal data when the variable is repeatedly measured over time.

Second, statistical checks are powerful tools because the multivariate structure and/or time dependence of variables are very sensitive to deviations (in the case of errors) and hard to mimic (in the case of fraud) [11,12]. Fabricated or falsified data, even if plausible univariately, are likely to exhibit abnormal multivariate patterns that are detectable statistically. In addition, humans are poor random number generators, and are generally forgetful of natural constraints in the data. Tests on randomness can be used to detect invented data. Benford's law on the distribution of the first digits, or tests for digit preference, can raise red flags [17,18]. Tests on dates can also be useful in detecting abnormalities in the distribution of days (e.g. a high proportion of visits during weekends may reveal data fabrication) [12,18].

Third, a key tenet of statistical monitoring is that every piece of information collected in the CRF during the conduct of the trial, and every variable coded in the clinical database is potentially indicative of data quality, not just those associated with a set of indicators predefined to reflect site performance in terms of data quality ("key risk indicators"). A statistical approach therefore requires a large number of statistical tests to be performed: tests on proportions of outliers, means, global variances, within-patient variances, event counts, distributions of categorical variables, proportion of week days, proportion of missing values, correlations between several variables, and so on and so forth. These tests generate a high-dimensional matrix of p-values, which can be analyzed by statistical methods and bio-informatic tools to identify outlying centers.

CHALLENGES IN ACTUAL TRIALS

Although the principles of statistical monitoring are simple enough, its application to ongoing trials turns out to be quite challenging for the following reasons:

1. **Staggered availability of data**: the staggered opening of centers and the staggered accrual of patients result in substantial differences, especially early on, between follow-up times and variables available for analysis at different centers. Yet central monitoring is potentially most valuable early in the trial, in order to remedy problems before it is too late to do so.

2. **Volume of data**: statistical analyses are only informative if the volume of data to be analyzed is reasonably large, which is not the case early after trial start. In some therapeutic areas and in rare indications, the number of patients per center can remain small even when the study is completed, which may limit the potential to identify problems in any individual center.

3. **Cleanliness of data**: the data collected in ongoing trials are subject to typing mistakes, mis-specified units, etc., that result in aberrant values. Such outliers are typically detected and corrected by data management, but statistical monitoring cannot wait for full data cleaning. Although trivial errors may sometimes be informative of the performance of centers in terms of data quality, they can also dilute true signals by introducing random noise.

4. **Systematic differences between centers**: substantial variability is often observed between centers in multi-centric trials, due to a host of factors such as socio-economic differences between patient populations recruited, ethnic and cultural differences between countries, etc. A typical example would be the expected differences in the distribution of the patient's height and weight in a trial involving Asian, European, and North American countries. The central monitoring system must be able to adjust for such systematic differences.

RESULTS

The examples shown in this section come from analyses of actual trial datasets. These examples are illustrative only and do not aim at providing a comprehensive list of problems that can occur and be detected by either on-site or central statistical

Table 1. Repeated measurements of vital signs for a particular patient in center X

Visit	Systolic blood pressure (mm Hg)	Diastolic blood pressure (mm Hg)	Respiratory rate (breath/min)
Baseline	130	80	16
Visit 1 – pretreatment	130	80	16
Visit 2 – pretreatment	130	80	16
Visit 3 – pretreatment	130	80	16
Visit 4 – pretreatment	130	80	16
Visit 5 – pretreatment	130	80	16
Visit 6 – pretreatment	120	80	16
Visit 7 – pretreatment	120	80	16
Visit 8 – pretreatment	120	80	16

monitoring. Some details about the examples were hidden or modified to preserve source data anonymity.

EXAMPLE 1 - DATA PROPAGATION

The first example comes from a multicenter randomized trial in ophthalmology that was used for marketing approval of a new drug. At each visit, the attending ophthalmologists were asked to measure the blood pressure and respiratory rate of the patients. Central statistical analysis revealed abnormally low within-patient variability for one of the participating centers.

Table 1 shows the measurements for a particular patient from the outlying center.

The blood pressure and respiratory rates were remarkably constant over the visits in center X, whereas some natural variation was expected and present in most other centers. Propagation of values between visits is a well-known phenomenon in clinical trials [11,12], especially for variables considered unimportant (such as, arguably, vital signs in ophthalmology). The center(s) in which data were simply propagated over visits might well have escaped the scrutiny of on-site monitors who check source data against CRF. All values, taken individually, were plausible but the unusual pattern appeared clearly when taking the longitudinal dimension of the data into account.

EXAMPLE 2 – BIASED SCORES

The second example comes from a multicenter randomized trial for a psychiatric disease. The trial had two stages: an open label phase (eligibility screen), followed by a randomized double blind phase (for eligible patients only). Eligibility criteria included a psychiatric score, which had to be less than 20 for a patient to be eligible. Figure 1 shows the distribution of scores at entry in the open label phase (panel A) and prior to randomization (panel B).

Clearly, the shapes of these two distributions are different, with a clear tendency for investigators to report a lower score prior to randomization, so as to make the patient eligible for the trial (panel B). When

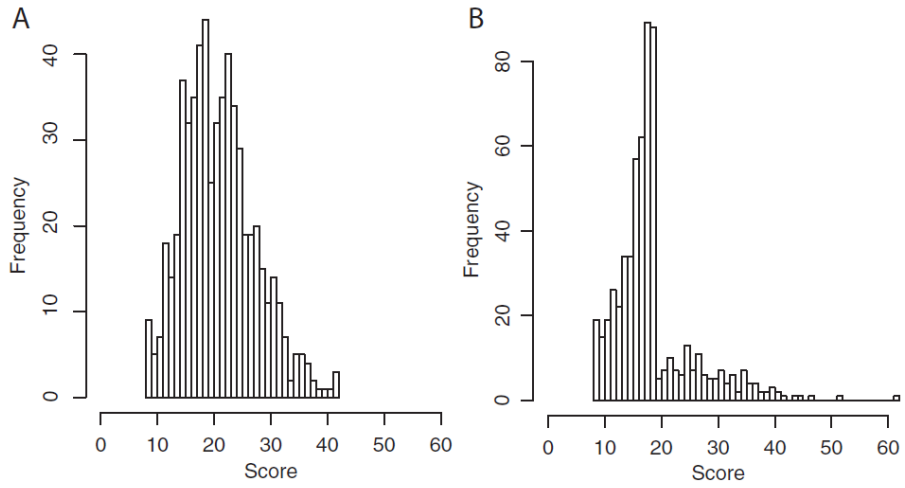

Figure 1. Distribution of psychiatric scores at entry in open-label phase (panel A) and prior to randomization (panel B). Patients with a psychiatric score below 20 were eligible.

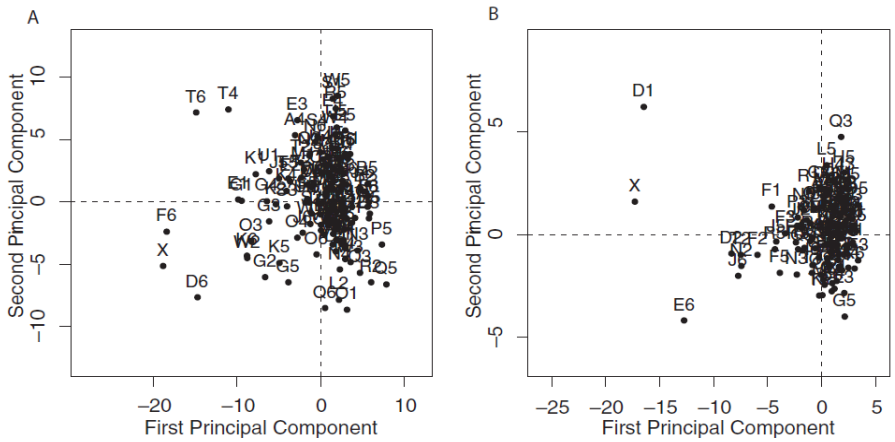

Figure 2. Principal component analysis of *p*-values. Centers are identified by a letter (denoting the country) and a number, except for center X where fraud had been identified. Centers far from the origin behave differently from the bulk of centers around the origin and are therefore more likely to be problematic. Panel A: analysis based on all data. Panel B: analysis based on patient-questionnaire data. PCA: Principal component analysis.

the trial protocol was developed, it was assumed that about half of the patients would satisfy the eligibility criteria; in reality, two thirds of the patients were within the allowed criteria.

EXAMPLE 3 – KNOWN FRAUD

The third example is different in so far as it was known beforehand that one center was suspicious in this multicenter randomized trial. . Indeed, center X was found to commit fraud after an audit revealed that the staff at this center completed information that was supposed to be completed by the patients. Interest therefore focused on whether central statistical monitoring could have detected this problem. The graph in panel A of Figure 2 shows the two first dimensions of a principal component analysis performed on the matrix of log p-values of tests comparing the means of all variables.
A principal component analysis projects each multi-dimensional point (in this case center) into a smaller set of dimensions (two dimensions in this case), so that as much of the variability as possible is explained [19].

The interpretation of this graph is quite intuitive, as many centers are clustered around the origin: these are probably non-problematic centers. Centers far from the origin are possibly problematic. Center X, which falls furthest from the origin, appears as most likely to be problematic in this analysis.
A detailed review of the tests for which the p-values were highly significant in this center indicated both a shift in mean and in variance for the data coming from the patient questionnaire. The graph in panel B of Figure 2 shows the first two principal components based only on the analysis of the patient questionnaire data. Center X appears as an outlier in this graph, too.

Of note, other centers (D6 and F6 in the all-data analysis, and D1 and E6 in the patient-questionnaire analysis) are almost as extreme as center X in the principal component analyses. Further central analyses would be useful to pinpoint the discrepancies detected in these centers. Confirmation of the potentially problematic nature of those centers could then be obtained by an on-site audit. In such a situation, the contribution of central statistical monitoring to classical monitoring lies in its ability to inform monitors of the nature and magnitude of possible problems, making monitoring visits more targeted and informed, and hopefully more efficient.

EXAMPLE 4 – MIS-CALIBRATED EQUIPMENT

The last example comes from an ongoing multicenter randomized trial in which the mean value of a variable

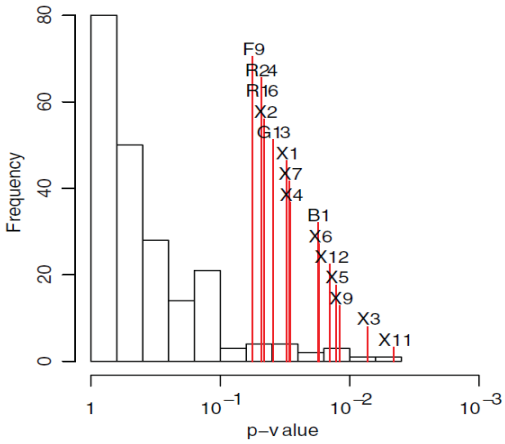

Figure 3. Frequency distribution of *p*-values obtained by comparing a feature mean across centers. Centers are identified by a letter (denoting the country) and a number. Vertical lines indicate the *p*-values of specific centers (the height of these lines is unimportant). Country X clearly shows a large number of outlying centers.

in several centers in country X was shown to differ significantly from the mean value in all other centers. Figure 3 shows the distribution of the p-values obtained by comparing the mean of each center to the means of all other centers.
The distribution clearly identifies a group of outlying centers belonging to country X. On-site inspections revealed that all these centers used equipment from the same lot, which was found to be mis-calibrated. Note that individual p-values were not very significant because many centers were subject to the same problem.

DISCUSSION

In this paper we build on previous work that paved the way to central statistical monitoring [11-12, 18]. We focus on the detection of data errors and show, through actual examples, that central statistical monitoring can reveal data issues that had remained undiscovered after careful SDV and on-site checks. These data issues may in turn point to other problems, such as lack of resources or poor training at the centers concerned, which would call for corrective actions. Central statistical monitoring allowed highlighting problems, such as lack of variability in blood pressure measurements or implausible values in a questionnaire, which would not have been detected by key risk indicator methods. This is because the former approach compares centers on all possible features, while the latter approach focuses on specific, pre-defined features thought to be of particular relevance to data quality.

Central statistical monitoring has limitations of its own. At the beginning of a trial, the amount of information available to perform statistical tests may not be sufficient to detect abnormal trends or patterns in the data. In trials with many small centers, statistical monitoring may fail to detect centers with serious data problems, again due to the lack of sufficient information for statistical tests to raise red flags. Statistical monitoring relies on computerized data only, and may therefore miss some types of fraud or serious errors that can be detected during site visits or audits (such as evidence provided in hand-written documents, or in interviews with site personnel). Conversely, on-site visits may be made more efficient if monitors perform these visits with information about unexpected or strange data patterns identified by statistical monitoring in hand. Targeted monitoring differs from central statistical monitoring in that it relies on “key risk indicators”, the drawbacks of such an approach being the programming required for every new study, and the fact that not all data are exploited. In contrast, statistical monitoring takes advantage of all the data and requires no trial-specific programming, at the expense of being less specific.

There is currently little empirical evidence that reduced or targeted monitoring strategies may achieve the same level of data quality than extensive monitoring with full source data verification – in fact, there has also been little theoretical research into quality standards for clinical trials [8]. There is, however, some evidence that the findings made during on-site visits can be detected by central monitoring. Investigators at the UK Medical Research Council recently reviewed the findings made during monitoring visits in a large trial conducted in patients with HIV (human immune-deficiency syndrome) in Africa. Of 268 monitoring findings, 76 (28%) were also identified in the central database, 179 (67%) could have been identified through central checks, had these been in place, and only 13 (5%) would have required a site visit to be found [20]. Clearly, extensive data checks during on-site monitoring visits are neither cost-effective nor sustainable; in contrast, a statistical approach to quality assurance can yield large cost savings and yet increase the reliability of trial results.

CONCLUSION

Central monitoring through advanced statistical and bio-informatic methods can detect abnormal patterns in the data. It can also help improve the effectiveness of on-site monitoring by prioritizing site visits and by guiding site visits with central statistical data checks.

FUNDING

The first author was supported by a FIRST post-doctoral grant of the Walloon Region, Belgium.

ACKNOWLEDGMENTS

The contents of this paper were presented in part by the third author at the Conference on “Developing effective quality systems in clinical trials: an enlightened approach”, Washington DC, October 13-14, 2010 and by the last author at the University of Pennsylvania Annual Conference on Statistical Issues in Clinical Trials, Philadelphia, PA, April 13, 2011.

REFERENCES

1. **Eisenstein E, Collins R, Cracknell,** et al. Sensible appoaches for reducing clinical trial costs. Clin Trials 2008; 5: 75-84.
2. **Califf, RM**. ACS and Acute Heart Failure Models. Speaker presentation at the Institute of Medicine Workshop on Transforming Clinical Research in the United States, October 7–8, 2009, Washington DC. Available at http://www.iom.edu/Reports/2010/Transforming-Clinical-Research-in-the-United-States.aspx (accessed 13 November 2011).
3. U.S. Department of Health and Human Services, Food and Drug Administration. Draft Guidance: Guidance for IndustryOversight of Clinical Investigations — A Risk-Based Approach to Monitoring. Available at http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM269919.pdf (accessed 13 November 2011).
4. European Medicines Agency. Reflection paper on risk based quality management in clinical trials. EMA/INS/GCP/394194/2011. Available at http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2011/08/WC500110059.pdf (accessed 13 November 2011).
5. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH Harmonised Tripartite Guideline: Guideline for Good Clinical Practice E6(R1). Available at: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E6_R1/Step4/E6_R1__Guideline.pdf (accessed 13 November 2011).
6. **Baigent C, Harrel F, Buyse M,** *et al*. Ensuring trial validity by data quality assurance and diversification of monitoring methods. Clin Trials 2008; 5: 49-55.
7. **Morrison B, Cochran C, White J**, *et al*. Monitoring the quality of conduct of clinical trials: a survey of current practices. Clin Trials 2011; 8: 342.
8. **Grimes D, Hubacher D, Nanda K**, *et al*. The Good Clinical Practice guideline: a bronze standard for clinical research. Lancet 2005; 366: 172–74.
9. **Lyle K, Dent L, Bailly S,** *et al.* Carbon cost of pragmatic randomised controlled trials: retrospective analysis of sample of trials. Brit Med J 2009; 339: b4187
10. **Cooley S, Srinivasan**. Triggered Monitoring. Applied Clinical Trials. August 2010. Available at http://appliedclinicaltrialsonline.findpharma.com/appliedclinicaltrials/CRO%2FSponsor/Triggered-Monitoring/ArticleStandard/Article/detail/682214 (accessed 13 November 2011).
11. **Buyse M, George S, Evans S**, *et al*. The role of biostatistics in detection and treatment of fraud in clinical trials. Statist Med. 1999; 18: 3435-3451.
12. **Evans S**. Statistical aspects of the detection of fraud. **In**: Lock S, Wells F, Farthing M (Eds). *Fraud and Misconduct in Biomedical Research*. London: BMJ Publishing Group, 2001 (third edition).
13. **Lienard J-L, Quinaux E, Fabre-Guillevin E**, *et al*. Impact of on-site initiation visits on patient recruitment and data quality in a randomized trial of adjuvant chemotherapy for breast cancer. Clin Trials 2006; 3: 486-492
14. **Chêne G, Jounot V**. OPTIMON study: OPTImisation of MONitoring. Available at https://ssl2.isped.u-bordeaux2.fr/OPTIMON/Documents.aspx (accessed 13 November 2011).
15. **Brosteanu O, Houben P, Ihrig K,** *et al*. Risk analysis and risk adapted on-site monitoring in noncommercial clinical trials. Clin Trials 2009; 6: 585-596
16. **The ESPS2 Group**. European Stroke Prevention Study 2. Efficacy and safety data, J Neurol Sciences 1997; 151 (Suppl): S1-S77.
17. **Hill TP**. A statistical derivation of the significant-digit law. Statistics in Science 1996; 10: 354-363.
18. **Buyse M, Evans S**. Fraud in clinical trials. In: Biostatistics in Clinical Trials (Colton T, Redmond C, eds.), Wiley, New York, 2001.
19. **Härdle W, Simar L**. Applied Multivariate Statistical Analysis. Springer, New York, 2003.
20. **Bakobaki JM, Rauchenberger M, Joffe N**, *et al*. The potential for central monitoring techniques to replace on-site monitoring: findings from an international multi-centre clinical trial. Clin Trials 2012; 9: 257-64.